Signal peptides of secreted proteins of the archaeon Sulfolobus solfataricus: a genomic survey

Albers, S.V.; Driessen, A.J.M.

Published in:
Archives of Microbiology

DOI:
10.1007/s00203-001-0386-y

Arch Microbiol (2002) 177 : 209–216
DOI 10.1007/s00203-001-0386-y

MINI-REVIEW

Sonja-Verena Albers · Arnold J. M. Driessen

Signal peptides of secreted proteins of the archaeon *Sulfolobus solfataricus*: a genomic survey

Received: 21 August 2001 / Revised: 31 October 2001 / Accepted: 15 November 2001 / Published online: 11 January 2002

Abstract Analysis of the recently completed genome sequence of the thermoacidophilic archaeon *Sulfolobus solfataricus* reveals that about 4.2% of its proteome consists of putative secretory proteins with signal peptides. This includes members of the four major classes of signal peptides: secretory signal peptides, twin-arginine signal peptides, possible lipoprotein precursors, and type IV pilin signal peptides. The latter group is surprisingly large compared to the size of the groups in other organisms and seems to be used predominately for a subset of extracellular substrate-binding proteins.

Keywords *Sulfolobus solfataricus* · Archaea · Signal peptide · Secretion · Type IV pilin signal peptide · Sugar-binding protein

Introduction

Sulfolobus solfataricus is an obligate aerobic archaeon that grows either litoautotrophically or chemoheterotrophically in hot (about 80 °C) and acidic (pH 2–4) environments. *S. solfataricus* P2 originates from a solfataric field near Naples, Italy (Zillig et al. 1980), and its genome sequence has recently been determined (She et al. 2001). As a model organism for the domain of crenarchaeotes, its mechanisms of cell cycle, DNA replication, chromosomal integration, transcription and translation have been studied extensively. Furthermore its membrane-spanning tetraether lipids, metabolic routes and sugar degradation pathways are unique (Schönheit et al. 1995). Only limited data are available on the secreted proteins and secretory apparatus of *S. solfataricus*. Here, we briefly discuss bacterial and eukaryal secretion mechanisms and substrates and use this information to classify the identified and putative secreted proteins of *S. solfataricus* that are present in its proteome.

In gram-negative bacteria, the general secretion system directs proteins to the periplasmic space and the outer membrane (Pugsley 1993). Various other secretion mechanisms are involved in the delivery of macromolecules to the extracellular medium, a process that involves a translocation step across the outer membrane. These systems are classified in four groups (Nunn 1999). Type I secretion systems consists of three proteins including an ATPase that belongs to the ABC-type of transporters. This system mediates the translocation of proteins across the inner and outer membrane without the accumulation of a periplasmic intermediate. The type II secretion system is formed by over 12 subunits that reside in the inner and outer membrane. This system handles the translocation of periplasmic substrate intermediates prior to their translocation across the outer membrane (Sandkvist 2001). Type III secretion systems are involved in eukaryotic host invasion mechanisms and are complicated structures with up to 30 subunits. These systems include an injection device that delivers macromolecules directly from the bacterial cytoplasm into the host cells (Tamano et al. 2000). Type IV secretion systems are involved in various processes such as single-stranded DNA transport into host cells (the Vir system of *Agrobacterium tumefaciens*), toxin secretion and pilin biogenesis (Christie 2001). Some of the components of the type IV secretion machinery, in particular the subunits of the macromolecule-conducting pore in the outer membrane, exhibit striking similarities with type II secretion systems. Also subunits of the type IV pilin biogenesis apparatus that are involved in twitching motility share homology with components of the type II secretome (Wall et al. 1999). Secretion of proteins across the S-layer of archaea has barely been addressed experimentally. Although the S-layer contains pores with a diameter of 4–5 nm, it is not known if these structures conduct protein movements in an active or passive sense. In-

S.-V. Albers · A.J.M. Driessen (✉)
Department of Microbiology,
Groningen Biomolecular Sciences and Biotechnology Institute,
University of Groningen,
Kerklaan 30, 9751 NN Haren, The Netherlands
e-mail: a.j.m.driessen@biol.rug.nl,
Tel.: +31-5036-32164, Fax: +31-5036-32154

terestingly, archaea contain homologues of PilT, the motor protein of the type IV pilin biogenesis apparatus, and of the VirB proteins involved in type IV secretion.

Structure and function of signal peptides

Proteins translocated across the cytoplasmic membrane of bacteria, the thylakoid membrane in plant chloroplasts, and the endoplasmic reticulum (ER) membrane of eukaryotes are all synthesized as precursors with an amino-terminal signal peptide. These signal peptides are functionally exchangeable between the different organisms (von Heijne 1990), and although their amino acid composition show little similarity, three different domains can be distinguished (von Heijne 1990) (Pugsley 1993). The N-domain contains basic amino acid residues. In bacteria, the net positive charge of this domain is thought to orient the N-terminus in the cytoplasm according to the $\Delta\psi$, which in most organisms is inside negative (Andersson et al. 1994). The N-domain also interacts electrostatically with negatively charged phospholipids in the lipid bilayer during translocation (de Vrije et al. 1990) and with the translocation ATPase SecA. The H-domain is a stretch of about 10–15 hydrophobic residues that tends to fold into α-helical conformation in the membrane. A glycine or proline residue is often found in the middle of the H-domain and has been proposed to promote the insertion of the signal peptide into the membrane by forming a hairpin-like structure. Unlooping of this hairpin may result in the insertion of the complete signal peptide (de Vrije et al. 1990). The H-domain is followed by the short polar C-domain, which contains the recognition site for the signal peptidase. Recent studies have shown that the composition of the C-domain determines the accuracy of cleavage by type I signal peptidases (SPase), and not the length or even the presence of the H-domain (Carlos et al. 2000). After proteolytic cleavage by a signal peptidase, the mature protein is released from the membrane for further folding and assembly. The signal peptide is degraded by signal peptide peptidases.

Amino-terminal bacterial signal peptides can be divided into at least four different classes dependent on the signal peptidase recognition site (Fig. 1). Class 1 consists of the typical signal peptides, which are mostly cleaved by the type I signal peptidases (SPases) (Tjalsma et al. 2000). A subclass of these signal peptides contains a "twin-arginine" motif, which directs these proteins to a different translocation pathway, the Tat pathway (Berks et al. 2000). This pathway is mostly involved in the translocation of folded redox proteins with bound co-factor. Class 2 signal peptides exhibit a typical domain with an invariant cysteine that is lipid-modified prior to the cleavage by type II SPases. The resulting lipoprotein remains anchored to the cytoplasmic membrane after cleavage. Class 3 signal peptides include the type IV pilin-like peptides. Prepilins are cleaved between the N- and the H-domain, leaving the H-domain attached to the mature pilin. The remaining H-domain subsequently functions as a scaffold for the assembly of the subunits into the pilin structure.

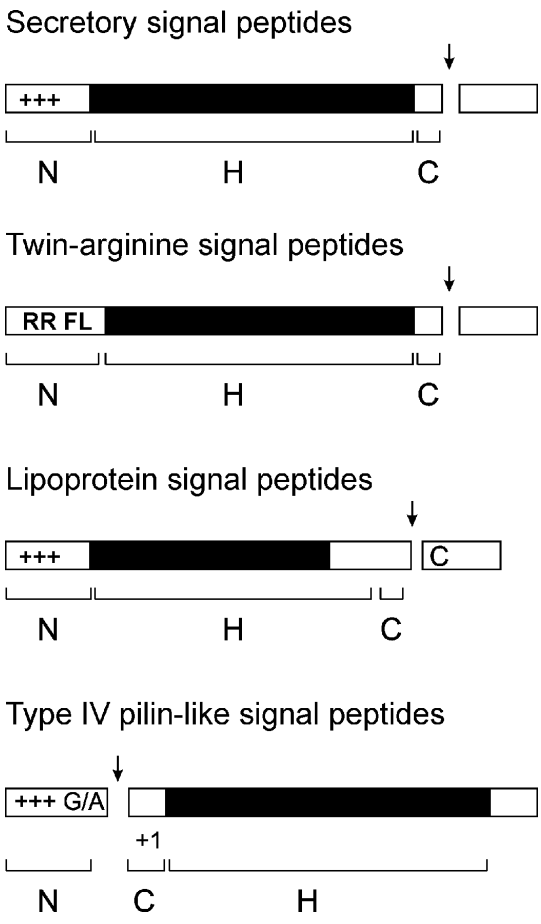


Fig. 1 The different classes of signal peptides found in *Sulfolobus solfataricus*. The length of the domains N, H and C is given and the arrow indicates the cleavage site. + Positive charges, *black box* hydrophobic residues

Class 4 signal peptides constitute a heterogeneous group of signal peptides such as the signals that direct the secretion of small antimicrobial peptides via ABC transporters.

Distribution of signal peptide classes in *Sulfolobus solfataricus*

This section describes the results of an analysis of the *S. solfataricus* genome which was screened for the distribution of the various classes of signal peptides. To identify and classify putative secreted proteins, the complete genome database (http://www-archbac.u-psud.fr/projects/sulfolobus/) was analyzed by a neural network-based method and a hidden Markov model (http://www.cbs.dtu.dk/services/SignalP/) trained on both eukaryotic and gram-positive and gram-negative bacterial signal peptide datasets. Polytopic membrane proteins were excluded from the analysis, but membrane proteins containing only one amino-terminal transmembrane segment may be falsely predicted as signal peptides. Signal peptides selected were screened for the presence of the twin-arginine, lipobox and type IV prepilin-like sequences and motifs.

ORF	Amino acid sequence	Function
SSO0011	MKVISVKKSLIILLFVILSPITYLTLPLSSQSTPIQGYATSSELITPGEI	hypothetical
SSO0012	MYMILELLNIIGIIAFTISGSLKGTNKGLDIFGVVTLGVITSYAGGIIAD	hypothetical
SSO0037	MKKGISSILGAIILIQIVVSSVGLILYLTSLNAKMSNIAYSQIYEELQNA	hypothetical
SSO0045	MIKVDRKEKFELYWVIYVIVLFAIVIGATAPAVYTVGGDLSSVQAGIIIP	terminal oxidase
SSO0055	MKRIIILSPFRGLFRSLLYFLLGLIMALISAGYFSQLFSIVGINRDIAII	hypothetical
SSO0117	MMWLKAISSIFSTLIVVMITLSLIVPLYLFFTQTYTNSSIQANSAYDNYL	hypothetical
SSO0118	MLQLMMKGGYKLKKRKGLSSILGTVIVLAITLVLGGLLYAYSNGLFSSLT	hypothetical
SSO0152	MGIENFTAKIVGLSIVSLLVVLMFLYKLIYIIPLIFIVLLVFQSEKKIFA	hypothetical
SSO0283	MKMNKNLWIPILLIILGIGFLFHNLININLMFFVFPIIMIVVISFIFRNS	hypothetical
SSO0309	MIVKIYPSKISGIIKAPQSKSLAIRLIFLSLFTRVYLHNLVLSEDVIDAI	EPSP synthase
SSO0330	MIRIALIGVGNVASALVQSIELIRNGKEIYGILDLPIRPNDIEIVAAFDI	hypothetical
SSO0335	MGKNFLNKFQLSSRSKMADMKTIAFSIVAVVLIVIAAIGFYEYSVANSRY	hypothetical
SSO0389	MNKTLGLILTSVFLLSTLGIITGFVIPTQAANSNDAAIYTIPSVTSVSNT	hypothetical
SSO0390	MVVKKTFVLSTLILISVVALVSTAVYTSGNVTFYSPSVNNQIYYVGKSVT	hypothetical
SSO0483	MISNLSDFLVVVVVFILLMAGDKNAGNTTKSIGRFLGEIRKRQNEFKNEL	hypothetical
SSO0497	MKALLAFIVLLLSLSALITSSFSIVIISPNIVKILSYAQVGNNIYSSPLW	hypothetical
SSO0519	MKWFLLLLLVFGVLGIIPITNGVITGPHPQFDSGGGFAGPFFTYSKTMTI	hypothetical
SSO0522	MKRHLLLVAPLFLLLISLNALAVTANQLGASTILTTYNSDNWAGVAYADE	hypothetical
SSO0537	MAKSIGIGSILIIISIIIIGSVATIFYLENVDVNISVNPIYWRIYSNNYE	hypothetical
SSO0538	MEEKRLSFFKWLGLALLFIVLPSAVAVVLSFSVPYYILHDMTLANALSTI	hypothetical
SSO0567	MKRASLILAFLIPVLVSSVIAAAQAPSDTAQGFAGINIGAGLAVGLAAIG	ATPase C chain
SSO0583	MSLKSYMQLVRIHNVIGAALGAIMGFLVSSQWYLELKGILLSALVVGLIA	ubiquitinase
SSO0647	MFIHMKSINKVAVIGAGVIGVGWTTLLLAKGYKVNLYTEKKETLEKALAK	dehydrogenase
SSO0650	MLLMNRQILLALALLVIVVMAIGVYEGNKYRTEISTVALGSQQTDGMYML	hypothetical
SSO0687	MSKIFSIITISLFLVSLLFIPLTSSATQSSFSASSQWLSSTPYVTPGERL	hypothetical
SSO0766	MNPKLTVTFLFLLLMVIMGNELQLENKILKGTTTVGIRVNDGVILAADRR	proteasome
SSO0775	MPKKYNRLYNEVINSYVILILIFILIIGILGVIAFPYYISPLNNGQALNS	hypothetical
SSO0809	MLMKILISGGAGFLGSHLTEALLEKGEEITIVDDLSTAKYFNIRKDVEFI	glucose dehydratase
SSO0810	MRIGVVGLGVVGLVTGAVLADQGHEVVGVDIDQNKVKGLQCNRSPIYEPG	glucose dehydrogenase
SSO0816	MNGFSSLLTCWKNYVAIIFASTLLSLLFSFLNLLISASILTLFYLIDLIV	hypothetical
SSO0840	MRLLLLLLMLTITLLSSVSSFTASVSVQYPKEVILGSKISINFSLTQQEI	hypothetical
SSO0898	MRIAILGGGVAGSTLAYLLSRINYEVTIFDINQHYVKPCGDIVPNIYTPP	geranyl hydrogenase
SSO0916	MKMKKSDIIIILFIALIYILMFSNIVQSASVEGVSMYPIFQNGALTFYVK	hypothetical
SSO0997	MANKKLFIWSNICSSMIYIFGSGLAGLSAAISLHKSGYKVTIISKKINGG	aspartate oxidase
SSO1027	MIIMIGKVIVVLAIILVGVFLLTHTNLFYHPQTPVSKGQEYYTTTNVQNI	hypothetical
SSO1053	MIKNSAFIALGIILIDILVIYFFLYMPFSLSTFYPSFLLGPIYNFNPIEY	hypothetical
SSO1079	MRNRLIIILLLLSLTLPIIPVNSQSTVVISSWGWGTPQNPIRVHPGYNDT	hypothetical
SSO1093	MKGRVRIRGIYATALTSIFSSLSYEIVQQSVEISERFMQEINNLSADITI	hypothetical
SSO1131	MASKSVLVIGAGPAGLSATKELANMGVNVVVVEREPFLGGTPKRLKYSLL	reductase
SSO1141	MYRYIFLMSMLLISIIPLVFASNPNMYQNPITLKEFREIGTLNANEEVIV	protease
SSO1167	MRGEEIIFIILIFLSFLNPLLTFSATSSLKYYSPSYLLLNNWKNQSIWIV	hypothetical
SSO1172	MIKIAILAMGNLPKTAKAFLTLFFLLSLISCSFLIPTSQSISVNFTVSSN	hypothetical
SSO1175	MYMKAKHLISLIVILTPLVTLLTSAVYTSGGITFYSPAYNGESYYTGQSI	hypothetical
SSO1262	MRFGLLTTIGFSLLVLSVLSINSPIQSLISISNPYSIAVPKIAYVTARLF	hypothetical
SSO1273	MYSVLSIKDKKIISLLILVATAISPIFAIAQSASSSPASTAITIISYNGN	hypothetical
SSO1288	MKREIYLLFAIFLTMLISISPIITSGYALITNFQTPVLSPALYRAYTPYV	hypothetical
SSO1297	MYKSIVLILLVLFPLMLSGFSNSSSTTPPFSYFITANWKTIPTLDNLTTI	hypothetical
SSO1320	MNAKIPIILTVIIVVSAFIVVFASTHSTTQENSTADAFHYTTLGEIHTYN	hypothetical
SSO1354	MNKLYIVLPVIVIIAIGVMGGIIYLHQQSLSVKPVTTTEFSTTTSTSTTT	endoglucanase
SSO1360	MNLGKIIGLVVFISLLIVQFAILNNPINLVQTSFNQVNSIIIPLFTSEPK	hypothetical
SSO1375	MSYRTLLSIIVIILVMISSFSILCIIISNAEITIN NNITDSNNIYMAPN	hypothetical
SSO1392	MRVIACFIRFVTLNVLMTISGITIKHFAYCPQIVRIESMGFTERVSEAMI	hypothetical
SSO1460	MKRYNYLIIAVLGILLVILSTFTTSLSILNPFLGIWYSSGNVKILNEIVS	penicillin acylase
SSO1464	MNKYILVSVLLILVTVSGVIGYLVGSTNHETTQTIVQGQKIIVYPQSDPA	hypothetical
SSO1573	MAKRIKGDVWSNLVLVATVLVYVVYIALAGYTLTHLPPIPSVVETENGTV	cytochrome *b* subunit
SSO1584	MRPKMLLLIPILLSLPTLALAANSSSPSLYVMYQQYSSTLQITPSHITYK	hypothetical
SSO1623	MSYFIFRKLDINNSYILFLLFLILLSLLVLRVTAIRLYLNLSHEEVYLMK	hypothetical
SSO1638	MKGYNYLMIGALVTLLVILSIFTTSLSILNPFLGIWYSSGNVNMLNEIVS	penicillin acylase
SSO1872	MIQEMKKEPRAVSNAVFAGVVIVLVIIAAVGFALYATKPSTAPSTVTTPP	hypothetical
SSO1873	MRRQTSRYLLITIVIIAIIVIAGSALSQLRSTSSPLIDKPIAGDVYNQLV	hypothetical
SSO1878	MLLMKVITLFLFLFIILIPIVNAGYDYGYEFGYVNMNGIQAIVTLYNISL	hypothetical

Fig. 2 Secretory signal peptides of *S. solfataricus*. Positive charged amino acids and hydrophobic residues are shaded *black* and *gray*, respectively. *ORF* Number of open reading frame, *BP* binding protein

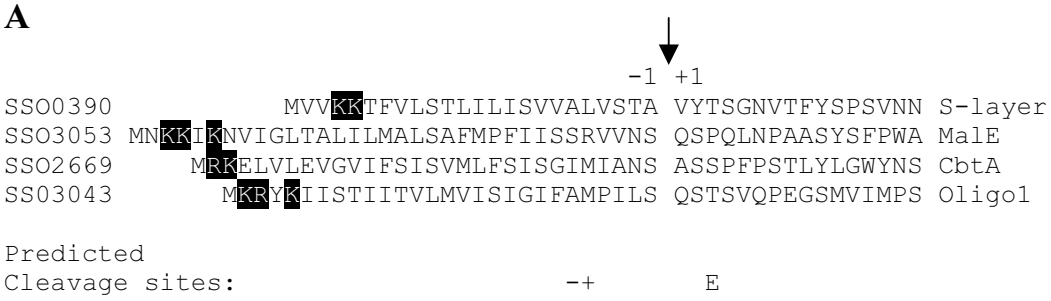
ORF	Amino acid sequence	Function
SSO1886	MLKHIVLVLLLLLLTPLVAISFPTGVVAYNGPICTNEVLGYANISSLLAY	thermopsin
SSO1933	MNAKKPIILTIIVLISAFIVVLASTHPTTQESGSTTDTFHYSTLDEIHTY	hypothetical
SSO1949	MIMNKLYIIIVPIIVIIVVGVIGGAIYLHHQSPNVKTSSITVTTNETTTL	endoglucanase
SSO1957	MDLGKIIGLIVFISVLIVQFAMINNFNGLLQTTLNQINSTILILFTSQPK	hypothetical
SSO2032	MHIHMDKKILIFIVLIVIISFALIMVYPSKDLLFSPQEAEAIFGGNWEVL	hypothetical
SSO2037	MQFRKTFLFLNIHFPYVLRNTLLILLLLLLPTPLLAISLPTGVVAYDGPIF	thermopsin
SSO2045	MRLLKILLLAMLILPLFSFFTLSISLYDQIQLPPHYLFYISENATQGSGI	hypothetical
SSO2050	MKSVEIVMKFSFLLLLIITVISKTFMLGNYIIHVNVNENFPLVQSVENST	hypothetical
SSO2067	MIELSTTKRLILGNEAIAFGALSAGVSVAAGYPGTPSTEIIETLMKFGKI	oxidoreductase
SSO2083	MAIGKTVLIVGAIILIVGIALFFIGGYLASSGLIKIVNTLSTASPTTLQP	hypothetical
SSO2088	MESKNVILKRVMLLLVLILSTTTFLTIIAQSQAQYYYIQTSSPQYTIIPG	peptidase
SSO2152	MLNIYMRKGLSDSVTMMIVLLASVILAITVVSILFTYLGYFGSNYGYVKQ	hypothetical
SSO2181	MTWSIFLLILALSDIVLPLTITNINNQSITTLSPNYYLTVAIVFPPSNLT	peptidase
SSO2194	MMYKVLLIIILLLPLSMPLSIPTTSQPSALAFPSGVTSYPLNTIIYTDFV	thermopsine
SSO2195	MAISIGDIVGIVFLIIIIIIFIAMSFRVVREWERAVVLRLGRFLRVKGPG	hypothetical
SSO2279	MIYGVLCMRSVTISILALIITWGILGSLIITQANTTIVNTTTIPTSITYT	hypothetical
SSO2319	MGVSQVVAYVLIFFITISLGLIALEAYIKSQQLLLHAENLRQNMELNQLT	hypothetical
SSO2322	MININLPELQQLVESPLFILLISISIPLAAFFISFFKIVLPRITRPKNIQ	hypothetical
SSO2326	MIREIYKLLLVGVISFLIIVTVISRLYIVLVPIVLFSIYLINESRIPEIK	hypothetical
SSO2420	MIEPVLNLAIIFISLAVLVIILMKIFGKSTAKFAYSDKALQLQNKSQKKK	AAA-type ATPase
SSO2488	MGKHLNTVPWDLVVMNTSIIIAVIVILIIVVGIVAYLTLVHHPATISSTT	sulfocyanin
SSO2551	MESRIIQVVVISTFLVLSVLFPLLSLAYSTTSINPSYPQSNVISALPSNT	serin protease
SSO2552	MRKNIALILLFSILAGIIVVPISSSQTSSSISHPLILGNSVLNSGKIPYD	hypothetical
SSO2570	MMAKRKKSQQNENKLIYIPFVVLAVVIVFLVAFPYFSHSSSLITANANTP	hypothetical
SSO2611	MNKAILGIVIVVLVLAGGVYYGFYYLTTGVVNVYIQDPPTSQGVKIYLTI	hypothetical
SSO2619	MSSLKGLALLSIMLIGIILPSLFLLQTSAQTSLTISPPNSSILIDVSQTA	binding protein
SSO2669	MRKELVLEVGVIFSISVMLFSISGIMIANSASSPFPSTLYLGWYNSNVEA	cellobiose BP
SSO2683	MKGQASVIAMVVIFFLIIATIGLILYISVSYENLQKEYIQVSQILANKAK	hypothetical
SSO2684	MRGISEAITVVFLILVTLIAIAIVTIYYLHIVNANQYGLYQELKNYYIDS	hypothetical
SSO2801	MLKPFCEKMSIKRKSKYTLGVLLLASFLAIIMGLANVPMAQTSPQIPVYK	cytochrome b
SSO2812	MKWYQILIIVVAVILVISGIVILVHNTSQQQINVTKVVFPTQAQISSILG	hypothetical
SSO2893	MVMLSVLVTRRGLSLTVVSILLLNPVGRVLYREAYLGDAVLRRAFYLAAL	transposon
SSO2964	MNRRNLITAIIGIIVIATGIIVYANHFIASAQPIPAGKFVKISNIDLAPK	hypothetical
SSO2967	MKKGKIVVPIVVMVIAALVVLSSGVLTVVNPFISTSAVSRELGGSWSVQN	hypothetical
SSO2969	MEKARIFELSTIVFAIVILTVLGVFSDIYLNSINTGAYLSTTQRQDAIPI	quinol oxidase
SSO2972	MKAQSSLLPVIVGVLVAIVAVGVSVYAYYEYQVLSAPTSTATSTSTSTSS	sulfocyanin
SSO3043	MKRYKIISTIITVLMVISIGIFAMPILSQSTSVQPEGSMVIMPSPGIVWQ	oligopeptide BP
SSO3053	MNKKIKNVIGLTALILMALSAFMPFIISSRVVNSQSPQLNPAASYSFPWA	maltose BP
SSO3089	MNKKQLIKALSSYQLWLIVGLIVIILIGVGAAYIMLKQSQSSSIPSSTQT	hypothetical
SSO3095	MSEKRASRVLTIILIVLILSEFCNGILVKTSTMKNIFISLIVLEGKPIMV	hypothetical
SSO3099	MASPPTSPLTVFATILISASSSPTIHVYYQRYDSVYRASGINGPLSASTV	hypothetical
SSO3104	MNKLLLLGVLLSTILVGGVVIGEEISGSLGTISYNVTSPTIQTTLASFNL	hypothetical
SSO3138	MRKAQSEYIGFIIAIIIVLIVIPLFYILSNYSVPSAKQLDYVQVLKNQI	hypothetical
SSO3139	MGILWRTLPRWGVMIMNKGLSNVISVIILIFIILLVLVPMLYYLEYSSQY	hypothetical
SSO3140	MKKALSSAIFLIITLIILLSVLIPALLIFNSTPIYSSQGQIAGTGYQQLQ	hypothetical
SSO3142	MPSAVTNLLIIIATVIITLSAFAIYSTFLSVQGVTFLQEENVISISKTVQ	hypothetical
SSO3175	MKNKFIMYIILFLLLISLSISTIGNINIQKEQVIIQQYSTYLIQNGESLN	hypothetical
SSO3177	MKTSILALTLVGAFLAGLATAGVAGYGPLAYISYHIMVSQQKGQAQVIPA	hypothetical
SSO3181	MRRLSLLTLTLLFLTPLMSHGNVVIVNYSSYQIHGSEILYSYSNSSYLIQ	hypothetical
SSO5023	MIRAFLLLTLFFRPIYSELALVLSLILILLILLLSSLSLLIKILRLLKNI	hypothetical
SSO5098	MMNMVSFFKLLGIGYILAIALLVWELTEETLHAAATSYILPFTIGAFIGF	hypothetical
SSO6024	MTAKAVSPFVCPICLTPFSSSSALKQHIRYEEHGKECQICKKRFTTTDAT	SSV1 homologue
SSO6661	MLINYDITLLVAFSSANCVPLVSFKPNVANADIIPHNAKNVITIQEFAKY	hypothetical

Fig. 2 (continued)

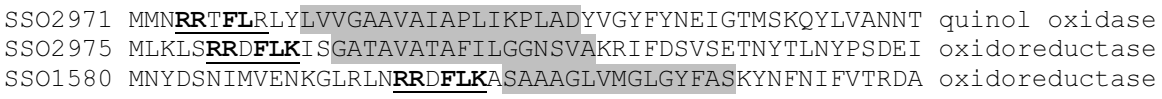
Class 1: secretory signal peptides/twin-arginine signal peptides

From *S. solfataricus* database, 114 proteins (about 3.9% of the proteome) are predicted to contain secretory signal peptides (Fig. 2). The presence of the signal peptidase cleavage site was recently confirmed by N-terminal amino acid sequencing of three sugar-binding proteins (Albers et al. 1999a; Elferink et al. 2001) and of the small (40 kDa) subunit of the S-layer (Fig. 3A). The N-domain is positively charged, with an average of two positively charged residues and a bias for arginine compared to lysine. The exact cleavage site is difficult to predict as the three neuronal networks yield different answers (eukarya,

A



B



C

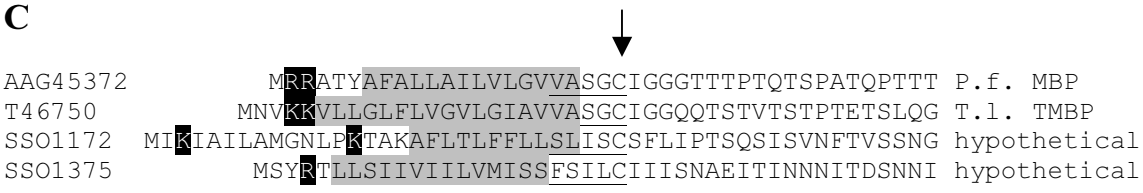


Fig. 3A–C Signal peptides of *S. solfataricus*. **A** Experimentally determined secretory signal peptides. Only the first 50 amino acids are shown. Prediction of potential cleavage sites for a known signal peptide dependent on matrix used for prediction (– gram-negative bacteria, + gram-positive bacteria, *E* eukaryotic) is only shown for SSO3043. **B** Putative twin-arginine signal peptide. The twin-arginine motif is underlined and residues according to the bacterial consensus are *bold*. **C** Putative lipoprotein signal peptides. For comparison the trehalose/maltose binding protein of *Thermococcus litoralis* and the maltotriose binding protein of *Pyrococcus furiosus* are shown. The "lipobox" is *underlined*. The positively charged residues of the N-domain are boxed in *black*. The putative H-domain is boxed in *gray*. *Arrows* indicate the cleavage site

gram-positive and gram-negative bacteria) (Fig. 3A). However, from the experimentally determined cleavage sites, it appears that the eukaryotic type of cleavage is preferred. In analogy to what has been reported for *Methanococcus jannaschii* (Nielsen et al. 1999), the H-domain of the archaeal signal sequences are equipped with a higher content of isoleucine and leucine residues than found in typical bacterial signal peptides. An unusual feature of the archaeal signal peptides is the clustering of bulky tyrosine residues around the putative signal-peptide cleavage site.

Three proteins are predicted as possible substrates of the Tat secretion pathway (Fig. 3B). These proteins, the quinol oxidase subunit SoxF (SSO2971), and the oxidoreductases (SSO2975 and SSO1580) exhibit a typical bacterial twin-arginine motif (kRRkFLK; see Berks et al. 2000) and may constitute ideal substrates for this pathway that appears to be involved mainly in the translocation of redox protein subunits that contain a cofactor. Several other signal peptides contain two arginine residues but they lack the remainder of the motif. These signal peptides therefore either follow the classical route or signify some degenerated twin-arginine motif. *S. solfataricus* contains two TatC homologues (SSO0484/SSO3108); TatC is thought to form the putative membrane pore of the Tat secretion machinery. Therefore, this mode of secretion also appears to be present in *S. solfataricus*.

Class 2: lipoprotein signal peptides

In bacteria, lipoprotein signal peptides are usually shorter than secretory signal peptides and contain a lipobox with the consensus sequence [I/L/G/A/]-[A/G/S]-C. The invariant cysteine at the +1 position of the mature lipoprotein is the site that is lipid-modified before the signal sequence is removed by SPase II. Lipoproteins have not yet been identified in archaea. However, for the halocyanin of *Nantronobacterium pharaonis*, a lipid modification has been suggested based on electron mass-spectrometry analysis (Mattar et al. 1994). Some archaeal solute-binding proteins contain a typical lipobox motif: SGC. This motif is, for instance, present in the trehalose/maltose binding proteins of *Thermococcus litoralis* and *Pyrococcus furiosus* (see Fig. 3C) (Horlacher et al. 1998; Koning et al. 2001). Since both proteins are N-terminally blocked, the cleavage site could not be directly determined and the presence of a lipid moiety can not be excluded. Few *S. solfataricus* signal peptides contain a cysteine residue and only two possible candidates could be identified that might be lipid-modified and processed by a lipoprotein signal peptidase (Fig. 3C).

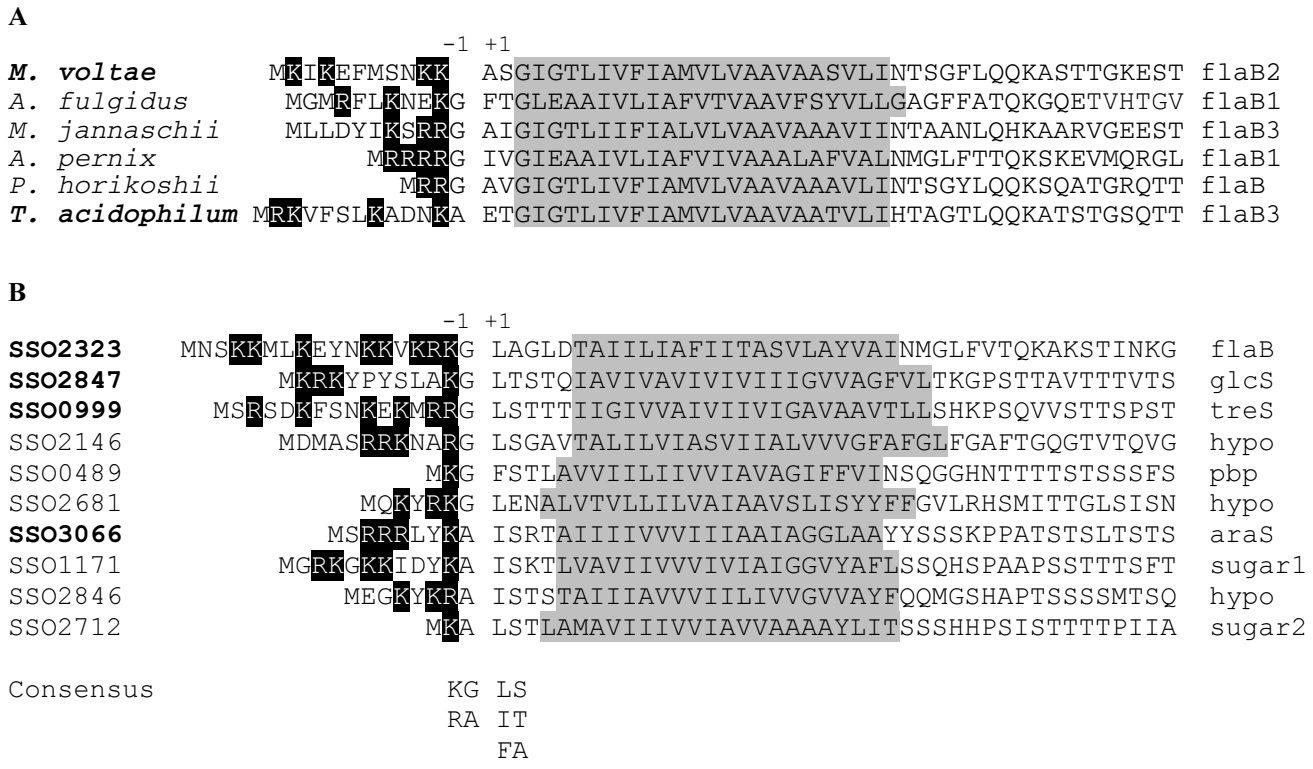


Fig. 4A, B Alignment of archaeal type IV pilin signal peptides. **A** Archaeal flagellins. The cleavage site was determined experimentally for the flagellins shown in *bold* (Thomas et al. 2001a). **B** *S. solfataricus* proteins exhibiting type IV pilin signal peptides. For the proteins displayed in *bold*, the N-terminus of the mature protein has been determined. Positive charges in the N-domain are *boxed*. The H-domain is *underlined*. *Hypo* Hypothetical protein, *pbp* putative phosphate-binding protein

Class 3: type IV pilin-like signal peptides

In bacterial prepilins, the processing site is located in between the N- and H-domains (Fig. 1). Since only the N-domain of the signal peptide is removed proteolytically, the H-domain remains attached to the mature pilin. Upon cleavage, the +1 residue is N-methylated. In bacteria, this residue is usually a phenylalanine. Faguy et al. (1994) first noted the occurrence of type IV pilin-like signal peptides in archaea by examining *Methanococcus voltae* flagellins. All archaeal flagellins exhibit a short, positively charged signal peptide of 4–18 residues (Fig. 4A). The –2 position contains a conserved positive charge (K/R), followed by a glycine at –1. The flagellin (flaB3) of *Thermoplasma acidophilum* seems somewhat unusual as it harbors an alanine residue at the –1 position (http://www.biochem.mpg.de/baumeister/genome/). In contrast to the bacterial sequences, the archaeal +1 position is quite variable, but contains mostly a small hydrophobic residue (alanine, isoleucine). Recently, this type of signal peptides was also reported for a subset of sugar-binding proteins of *S. solfataricus* (Albers et al. 1999b; Elferink et al. 2001). In total, ten proteins of *S. solfataricus* appear to carry a type IV pilin cleavage site (Fig. 4B). The site of processing was experimentally verified for four of these proteins (Albers et al. 1999b; Elferink et al. 2001). It is of interest to note that most of these proteins are involved in solute binding – only one protein encodes a preflagellin. This would indicate that the *S. solfataricus* type IV signal peptidase exhibits the same specificity as PilD, the type IV signal peptidase from *Pseudomonas aeruginosa* (Strom et al. 1994) (see also below). Remarkably, sugar-binding proteins that harbor this unusual type IV signal sequence are completely absent in the genome of *Sulfolobus tokodaii* (Kawarabayasi et al. 2001). Little information is available about the physiology of this organism, but since these proteins are lacking, one would predict that *S. tokodaii* is less versatile in its ability to utilize sugars than is *S. solfataricus*. The size of the *S. tokodaii* genome is about 2.7 Mb, which is almost 300 kb smaller than that of *S. solfataricus*.

Signal peptidases

Type I signal peptidases

Type I signal peptidase (SPases I) removes the signal peptides from secreted proteins at the *trans* site of the cytoplasmic membrane during or after translocation. This process is a prerequisite for the release of the mature protein (Dalbey et al. 1997). SPases can be divided into two classes: the P (prokaryotic)-type SPases, which are present in bacteria and organelles, and the ER-type SPases, which are present in the ER (Dalbey et al. 1997). The two classes mainly differ in the active site. Whereas in the P-type SPases a Ser-Lys catalytic dyad is involved in

cleavage (Paetzel et al. 1997), in ER-type SPases the lysine is replaced by a conserved histidine (Dalbey and von Heijne 1992; Van Dijl et al. 1992). The latter enzyme is thought to employ a Ser-His-Asp catalytic triad (Tjalsma et al. 2000). The SPases of archaea belong to the ER-type. Like yeast and bacteria, most archaea contain only one type I SPase. *S. solfataricus* has one typical ER-type SPase (SSO0916). In addition to the two transmembrane segments, it contains the conserved domains that are found also in the *Archaeoglobus fulgidus* Spc21 and *Bacillus subtilis* SipW (Tjalsma et al. 2000). Most eukaryotes contain two type I SPases (Dalbey et al. 1997), but the largest number of paralogous type I SPases have been identified in *A. fulgidus* (four proteins) and *B. subtilis* (seven proteins) (Tjalsma et al. 2000).

Type II signal peptidases

Type II SPases are required for the processing of lipid-modified preproteins. All known type II SPases are integral membrane proteins with four transmembrane segments (Munoa et al. 1991). As discussed before, there is only limited evidence that this type of lipid modification occurs in archaea (Mattar et al. 1994). Database searches yield no clear homologues of the bacterial type II SPase in *S. solfataricus* and other archaea. Thus, either this type of lipid modification does not exist in archaea or the archaeal enzyme is too distinct from the known type II SPase known to this date.

Type IV pilin peptidases

The best-characterized type IV pilin peptidase is PilD from *P. aeruginosa* (Strom et al. 1994). PilD is a bifunctional enzyme. At the cytoplasmic site of the membrane it cleaves the positively charged N-domain of the signal peptide of the prepilin and subsequently N-methylates the newly formed N-terminus of the mature pilin. The site of activity of PilD is the cytoplasmic site of the membrane. The presence of type IV pilin-like signal peptides implies the existence of a PilD homologue in archaea, but a candidate gene cannot be identified in *S. solfataricus* or other archaea by means of sequence similarity searching (Thomas et al. 2001a). In vitro assays for the processing of flagellins (preFlaB) of *M. voltae* (Correia et al. 2000) and binding proteins and flagellins in *S. solfataricus* (Albers and Driessen, unpublished observations) demonstrate that such an enzyme must be present. It has been suggested that the archaeal type IV signal peptidase that is responsible for the processing of the preflagellins is distinct from the enzyme that processes the binding proteins (Thomas et al. 2001a). This suggestion was based mainly on the observation that many of the known archaeal preflagellins are cleaved at a GA motif, whereas with binding proteins a GL or AI motif is prevalent. Strikingly, *S. solfataricus* preFlaB, however, harbors a GL motif (Fig. 4). Moreover, mutational studies with *M. voltae* preFlaB (Thomas et al. 2001a) and the *S. solfataricus* preGlcS (Albers and Driessen, unpublished observations) demonstrate that the archaeal peptidase is equipped with a specificity that is as broad as that observed for the bacterial enzyme. For the methanococcal preFlaB, it has been demonstrated that the positive charge at position –2 is absolutely required for cleavage (Thomas et al. 2001b). The consensus cleavage sequence for the archaeal type IV signal peptidase is [K/R][G/A]-[L/I/F]. It therefore seems most likely that a single peptidase is responsible for processing of both the flagellins and the binding proteins. Future studies should reveal the identity of this enzyme that, as far as sequence concerns, is not related to the bacterial peptidase.

Concluding remarks

Approximately 4.2% of the *S. solfataricus* proteome specifies putative secretory proteins with amino-terminal signal peptides, which is relatively low compared to the proteome of gram-positive *B. subtilis*, with nearly 7% (Tjalsma et al. 2000). This percentage is much higher than in *M. janaschii* (2%). Nonetheless, methanogen-like *M. janaschii* contains far less solute transporters than *S. solfataricus*, and a large number of the secreted proteins in the latter organism are binding proteins that are involved in the uptake of solutes via ABC-type transporters. The majority of the putative secreted *S. solfataricus* proteins have an unknown function. Only two are homologous to known extracellular degrading enzymes (endo-glucanases) and are probably released into the medium. Many of the putative secretory proteins (about 30%) contain a C-terminal hydrophobic sequence that may function as a transmembrane segment. This group of secreted proteins includes members of the di/oligopeptide binding proteins, proteases and the S-layer protein. Except for flagellin, the proteins with type IV pilin signal peptides are most likely membrane-bound via their N-terminal transmembrane segment. However, in analogy to the flagellins, this domain may also be involved in an assembly event. Secretion and assembly of flagellins into the flagellum in archaea most likely involves a type II secretion system, because some of the proteins present in the flagellin operon share homologies with bacterial type II secretion systems. Moreover, studies on bacterial type IV pili suggest that this secretion system is in many aspects structurally and mechanistically similar to type II secretion systems. Therefore, the study of the secretion of flagellins and binding proteins will provide important insight into protein secretion in the third domain of life.

Acknowledgements We thank Jelto Swaving for critical reading of the manuscript. This work was supported by a TMR grant of the European Commission (ERBFMBIC971980) to S.-V. Albers.

References

Albers S-V, Elferink MGL, Charlebois RL, Sensen CW, Driessen AJM, Konings WN (1999a) Glucose transport in the extremely thermoacidophilic *Sulfolobus solfataricus* involves a high-affinity membrane-integrated binding protein. J Bacteriol 181: 4285–4291

Albers S-V, Konings WN, Driessen AJM (1999b) A unique short signal sequence in membrane anchored proteins of Archaea. Mol Microbiol 31:1595–1596

Andersson H, von Heijne G (1994) Membrane protein topology: effects of $\Delta\mu$ H^+ on the translocation of charged residues explain the 'positive inside' rule. EMBO J 13:2267–2272

Berks BC, Sargent F, Palmer T (2000) The Tat protein export pathway. Mol Microbiol 35:260–274

Carlos JL, Paetzel M, Brubaker G, et al (2000) The role of the membrane-spanning domain of type I signal peptidases in substrate cleavage site selection. J Biol Chem 275:38813–38822

Christie PJ (2001) Type IV secretion: intercellular transfer of macromolecules by systems ancestrally related to conjugation machines. Mol Microbiol 40:294–305

Correia JD, Jarrell KF (2000) Posttranslational processing of *Methanococcus voltae* preflagellin by preflagellin peptidases of *M. voltae*. J Bacteriol 182:855–858

Dalbey RE, von Heijne G (1992) Signal peptidases in prokaryotes and eukaryotes – a new protease family. Trends Biochem Sci 17:474–478

Dalbey RE, Lively MO, Bron S, van Dijl JM (1997) The chemistry and enzymology of the type I signal peptidases. Protein Sci 6:1129–1138

De Vrije GJ, Batenburg AM, Killian JA, de Kruijff B (1990) Lipid involvement in protein translocation in *Escherichia coli*. Mol Microbiol 4:143–150

Elferink MGL, Albers S-V, Konings WN, Driessen AJM (2001) Sugar transport in *Sulfolobus solfataricus* is mediated by two families of binding protein-dependent ABC transporters. Mol Microbiol 39:1494–1503

Faguy DM, Jarrell KF, Kuzio J, Kalmokoff ML. (1994) Molecular analysis of archael flagellins: similarity to the type IV pilin-transport superfamily widespread in bacteria. Can J Microbiol. 40: 67–71

Horlacher R, Xavier KB, Santos H, DiRuggiero J, Kossmann M, Boos W (1998) Archaeal binding protein-dependent ABC transporter: molecular and biochemical analysis of the trehalose/maltose transport system of the hyperthermophilic archaeon *Thermococcus litoralis*. J Bacteriol 180:680–689

Kawarabayasi Y, Hino Y, Horikawa H, et al (2001) Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain7. DNA Res 8: 123–140

Koning SM, Elferink MGL, Konings WN, Driessen AJM (2001) Cellobiose uptake in the hyperthermophilic archaeon *Pyrococcus furiosus* is mediated by an inducible, high affinity ABC-transporter. J Bacteriol 180:680–689

Mattar S, Scharf B, Kent SB, Rodewald K, Oesterhelt D, Engelhard M (1994) The primary structure of halocyanin, an archaeal blue copper protein, predicts a lipid anchor for membrane fixation. J Biol Chem 269:14939–14945

Munoa FJ, Miller KW, Beers R, Graham M, Wu HC (1991) Membrane topology of *Escherichia coli* prolipoprotein signal peptidase (signal peptidase II). J Biol Chem 266:17667–17672

Nielsen H, Brunak S, von Heijne G (1999) Machine learning approaches for the prediction of signal peptides and other protein sorting signals. Protein Eng 12:3–9

Nunn D (1999) Bacterial type II protein export and pilus biogenesis: more than just homologies? Trends Cell Biol 9:402–408

Paetzel M, Dalbey RE (1997) Catalytic hydroxyl/amine dyads within serine proteases. Trends Biochem Sci 22:28–31

Pugsley AP (1993) The complete general secretory pathway in gram-negative bacteria. Microbiol Rev 57:50–108

Sandkvist M (2001) Biology of type II secretion. Mol Microbiol 40:271–283

Schönheit P, Schaefer T (1995) Metabolism of hyperthermophiles. World J Microbiol Biotechnol 11:26–57

She Q, Singh RK, Confalonieri F, et al (2001) The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. Proc Natl Acad Sci USA 98:7835–7840

Strom MS, Nunn DN, Lory S (1994) Posttranslational processing of type IV prepilin and homologs by PilD of *Pseudomonas aeruginosa*. Methods Enzymol 235:527–540

Tamano K, Aizawa S, Katayama E, et al (2000) Supramolecular structure of the *Shigella* type III secretion machinery: the needle part is changeable in length and essential for delivery of effectors. EMBO J 19:3876–3887

Thomas NA, Bardy SL, Jarrell KF (2001a) The archaeal flagellum: a different kind of prokaryotic motility structure. FEMS Microbiol Rev 25:147–174

Thomas NA, Chao ED, Jarrell KF (2001b) Identification of amino acids in the leader peptide of *Methanococcus voltae* preflagellin that are important in posttranslational processing. Arch Microbiol 175:263–269

Tjalsma H, Bolhuis A, Jongbloed JD, Bron S, van Dijl JM (2000) Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. Microbiol Mol Biol Rev 64:515–547

Van Dijl JM, de Jong A, Vehmaanpera J, Venema G, Bron S (1992) Signal peptidase I of *Bacillus subtilis*: patterns of conserved amino acids in prokaryotic and eukaryotic type I signal peptidases. EMBO J 11:2819–2828

Von Heijne G (1990) The signal peptide. J Membr Biol 115:195–201

Wall D, Kaiser D (1999) Type IV pili and cell motility. Mol Microbiol 32:1–10

Zillig W, Stetter KO, Wunderl S, Schulz W, Priess H, Scholz I (1980) The *Sulfolobus* – "*Caldariella*" group: taxonomy on the basis of the structure of DNA-dependent RNA polymerases. Arch Microbiol 125:259–269